

mn

Notice of Allowability

Application No.

10/733,541

Examiner

Cheyne D. Ly

Applicant(s)

CAO ET AL.

Art Unit

2168

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to June 05, 2007.
2. ☒ The allowed claim(s) is/are 1-3,5-7 and 25-34.
3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some* c) ☐ None of the:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

* Certified copies not received: _____.

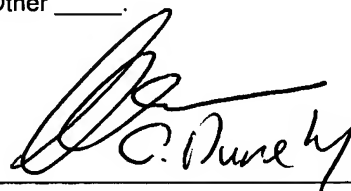
Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.

THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
- (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
- 1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____.
- (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☒ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application
6. ☒ Interview Summary (PTO-413), Paper No./Mail Date 8/13/2007
7. ☒ Examiner's Amendment/Comment
8. ☐ Examiner's Statement of Reasons for Allowance
9. ☐ Other _____

 8/18/07

Art Unit: 2168

EXAMINER'S AMENDMENT

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

1. Authorization for this examiner's amendment was given in a telephone interview with Todd Fronek on August 13, 2007.
2. The application has been amended as follows:

IN THE CLAIMS

REPLACE Claim 1 with claim 1 amended by examiner (without underlined and cross marked) set forth below:

1. A computer-implemented method of extracting information from an information source comprising a plurality of documents, comprising:
generating generalized extraction patterns, wherein the generalized extraction patterns express elements of consecutive patterns containing a wildcard, wherein the consecutive patterns specify a number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern;
accessing strings of text in the information source;
comparing the strings of text in the information source to the generalized extraction patterns and identifying a plurality of strings in the information source that match at least one generalized extraction pattern, the generalized extraction patterns including

Art Unit: 2168

related elements pertaining to a subject, at least one word and at least one wildcard, wherein the at least one word and at least one wildcard are positioned between the related elements and wherein the at least one wildcard denotes that at least one word and up to the specified number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern; extracting a first set of related elements of text pertaining to the subject from a first string of the plurality of strings based on the related elements pertaining to the subject in the at least one generalized extraction pattern, the first string being associated with a first document in the plurality of documents; extracting a second set of related elements of text pertaining to the subject from a second string of the plurality of strings based on the related elements in the at least one generalized extraction pattern, the second string being associated with a second document in the plurality of documents, wherein at least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of related elements of text; and outputting the first set of related elements and the second set of related elements.

REPLACE Claim 5 with claim 5 amended by examiner (without underlined and cross marked) set forth below:

5. A computer-readable storage medium for extracting information from an information source comprising a plurality of documents, comprising:

Art Unit: 2168

a data structure including a set of generalized extraction patterns, wherein the generalized extraction patterns express elements of consecutive patterns containing a wildcard, wherein the consecutive patterns specify a number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern, further, including related elements pertaining to a subject, at least one word and at least one wildcard, wherein the at least one word and at least one wildcard are positioned between the related elements and wherein the at least one wildcard denotes that the at least one word and up to the specified number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern; and

an extraction module using the set of generalized extraction patterns to match a first string and a second string in the information source with one of the generalized extraction patterns, the first string associated with a first document in the plurality of documents and the second string associated with a second document in the plurality of documents, extract a first set of related elements of text pertaining to the subject from the first string based on the related elements in said one of the generalized extraction patterns and a second set of related elements of text pertaining to the subject from the second string based on the related elements in said one of the generalized extraction patterns, wherein at least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of

Art Unit: 2168

related elements of text, and output the first of related elements and the second set of related elements.

The following is an examiner's statement of reasons for allowance:

1. The prior art of record fails to teach or suggest the claimed invention individually or in combination the limitations of "the generalized extraction patterns express elements of consecutive patterns containing a wildcard, wherein the consecutive patterns specify a number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern...at least one word and at least one wildcard, wherein the at least one word and at least one wildcard are positioned between the related elements and wherein the at least one wildcard denotes that at least one word and up to the specified number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern as set forth in claim 1, and similarly in claim 5.
2. Dependent claims 2, 3, 6-7, 25-34 being further limiting to the independent claim 1 or 5, respectively, definite, and enabled by the specification are also allowed.
3. Yangarber, closest prior art, discloses IE systems "for finding patterns automatically from un-annotated text" (page 282, Abstract etc.) as directed to pattern extractions. While, Soderland in view of Yanarber describes the performance of WHISK is comparable to other IE systems as directed to the application of wildcards to denote that at least one word in an individual string can be skipped in order to match the individual string to an individual generalized extraction

pattern. However, Yangarber in view of Sodeland does not teach or suggest the limitations cited above as being free of any prior art when read in the claims as a whole. Further, Muslea et al., provided with the instant Office Action, describes a method of extracting data from a document wherein a landmark is a sequence of tokens and wildcards (page 98, section 4. Extraction rules as finite automata. However, Muslea does not describe the landmark as the generalized extraction patterns express elements of consecutive patterns containing a wildcard, wherein the consecutive patterns specify a number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern.

4. Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

CONCLUSION

5. Patent applicants with problems or questions regarding electronic images that can be viewed in the Patent Application Information Retrieval system (PAIR) can now contact the USPTO's Patent Electronic Business Center (Patent EBC) for assistance. Representatives are available to answer your questions daily from 6 am to midnight (EST). The toll free number is (866) 217-9197. When calling please have your application serial or patent number, the type of document you are having an image problem with, the number of pages and the specific nature of the problem. The Patent Electronic Business Center will notify applicants of the resolution of the problem within 5-7 business days. Applicants can also check PAIR to confirm that the problem has been corrected. The USPTO's Patent Electronic Business Center is a complete

Art Unit: 2168

service center supporting all patent business on the Internet. The USPTO's PAIR system provides Internet-based access to patent application status and history information. It also enables applicants to view the scanned images of their own application file folder(s) as well as general patent information available to the public.

For all other customer support, please call the USPTO Call Center (UCC) at 800-786-9199. The USPTO's official fax number is 571-272-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to C. Dune Ly, whose telephone number is (571) 272-0716. The examiner can normally be reached on Monday-Friday from 8 A.M. to 4 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Tim Vo, can be reached on (571) 272-3642.

C. Dune Ly
Patent Examiner
8/18/07

